PDF RENDERIZATION OF Example2.R OUTPUT Bootstraped only 10 times to force equality of total effects

```
# Example 2: logistic models with multilevel categorical covariates
# Real data  from Spain's Encuesta Financiera de Familias (2002-2020)

library(OaxacaSurvey) # latest version of this package
library(fastDummies) # to transform multi-level categories into dummies
```

```
## Thank you for using fastDummies!

## To acknowledge our work, please cite the package:

## Kaplan, J. & Schlegel, B. (2023). fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from
```

```
library(data.table) # optional, for data import and handling
library(magrittr) # optional, for piping with %>%


# Import dataset with multilevel categorical data from a s
df <- fread("tests/eff-pool-2002-2020.csv")
df <- df[sv_year %in% c(2002, 2020)]
df[, group := 0][class == "worker" & sv_year == 2002, group := 0][class == "worker" & sv_year == 2020, g
df[, rentsbi := 0][rents >= renthog * 0.1 & rents > 2000, rentsbi := 1]
df[homeowner == "", homeowner := "Non-Owner"]
df$class <- relevel(as.factor(df$class), ref = "self-employed")
df$bage <- relevel(as.factor(df$bage), ref = "0-34")
df$inherit <- relevel(as.factor(df$inherit), ref = "Non-inherit")
df$homeowner <- relevel(as.factor(df$homeowner), ref = "Non-Owner")
df$riquezafin <- factor(as.logical(df$riquezafin), levels = c(TRUE, FALSE), labels = c("Fin", "NonFin"))
total_variables <- c(
  "facine3", "renthog", "renthog1", "bage", "homeowner", "worker", "young", "sex", "class",
  "actreales", "riquezanet", "riquezafin", "educ", "auton", "class",
  "tipo_auton", "direc", "multipr", "useprop", "inherit"
)
selected_variables <- c(
  "renthog", "bage", "sex", "homeowner", "riquezafin"
)

################################################################################

# select regressors
data <- df[, ..selected_variables]

# transform multi-level categories to dummies. IMPORTANT remove both first dummy and selected column to
data <- fastDummies::dummy_cols(data,
  select_columns = c("bage", "sex", "homeowner", "riquezafin"),
  remove_first_dummy = TRUE,
  remove_selected_columns = TRUE
)

# name the columns as x1....xn for oaxaca-blinder survey
colnames(data) <- paste0("x", seq_along(colnames(data)))
length_reg <- length(colnames(data))
new_formula <- paste("y ~", paste0("x", 1:length_reg, collapse = " + "))
```

```r
# compose the final dataset by adding: engenous variable, groups of split and sample weights
data <- cbind(y = df$rentsbi, group = df$group, weights = df$facine3, data)

# finally, test the model
result <- oaxaca_blinder_svy(
  as.formula(new_formula),
  data = data.frame(data),
  group = "group",
  weights = "weights",
  R = 10
)

# Return Oaxaca-Blinder decomposition with bootestraped CI
result %>% print()
```

```
##              unex          end        coef      inter        total   means1_y
## value  0.00893675 -0.014451843 -0.02169908 0.03570184  0.008487670 0.05457077
## CI1   -0.06933120 -0.018727449 -0.08484755 0.02753030 -0.004234032 0.04145344
## CI2    0.07980960 -0.007273721  0.06428280 0.04473007  0.017814909 0.06488637
##         means2_y    means_dif
## value 0.04608310  0.008487670
## CI1   0.04055121 -0.004234032
## CI2   0.05122521  0.017814909
```

```r
# plot the Oaxaca-Blinder decomposition in bars
result["value", ][1:4] %>%
  as.matrix() %>%
  barplot()
```